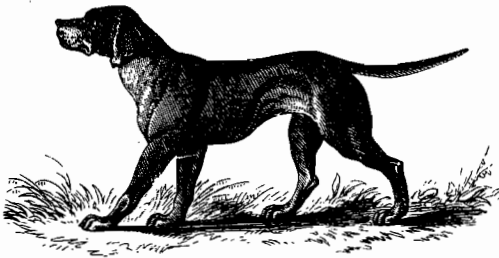


Animals, Mental Defectives, and the Social Contract

Tom Huffman
Teikyo Westmar University
LeMars, Iowa



Whatever its merit as an account of political institutions, the social contract theory has, in recent years, been the subject of intense scrutiny as an explanation for our *moral* obligations. Prominent examples of this approach include the work of John Rawls and David Gauthier.¹

A criticism of this understanding of moral obligation arises from several commentators, partly as a result of claims made by the original authors. The criticism is this: The social contract theory seems to imply that the scope of our moral obligations is unreasonably narrow. In particular, at first glance the theory seems to exclude from direct consideration the interests of animals and severely defective humans.

Criticisms of this sort are intuitive in nature and common to philosophy, yet many view them with disdain. It is an open question whether the failure of an otherwise attractive theoretical account to square with our moral intuitions tells against the theory or the intuition, but resolving this dispute lies beyond the scope of this paper. However, surely the intuition *and* the theory cannot both be true, or at least this is how it has seemed to the majority of commentators on the question. One either believes that we have direct duties to animals

and defective humans or that contractarianism is true, but one may not consistently believe both.

I wish to show two things. First, one may not rationally resolve this problem one way for defective humans and another way for animals. Regardless of one's substantive position on this dispute, whatever reasons one offers regarding one group will apply equally, *ceterius paribus*, to the other. Second, whatever methodological inclinations one holds toward intuitionism, one may consistently accept a contractarian account of morality and believe that we have direct duties to the parties in question. I will explain how this compatibilist position is possible, indeed quite plausible.

1. The Problem

The reason that contractarian accounts of morality appear inconsistent with possessing direct duties to beings at the margin of the moral community is fairly clear. If we understand moral obligations to be the set of rules that rational, self-interested beings would accept under a hypothetical set of circumstances, then such rules will be construed by and for rational creatures. Ordinarily, we enter into contracts for the purpose of bringing benefit, directly or indirectly, only to the contractors. If some beings are excluded *in principle* from participating in an agreement, and if the contractors lack a fundamental interest in their



PHILOSOPHY

welfare, then it is unclear how the terms of an agreement could be understood as directly benefitting them (although they may reap unintended, collateral benefit). Rawls writes:

Last of all, we should recall here the limits of a theory of justice. Not only are many aspects of morality left aside, but no account is given of right conduct in regard to animals and the rest of nature.... While I have not maintained that the capacity for a sense of justice is necessary in order to be owed duties of justice, it does seem that we are not required to give strict justice anyway to creatures lacking this capacity.²

Interpreting this passage is difficult insofar as Rawls goes on to say that it is nonetheless "wrong" to be cruel to animals because of "duties of compassion." Rawls seems to say that his is only a theory of justice and not a comprehensive theory of moral obligation and that we may have direct moral obligations to animals that go unexplained by his theory. However, insofar as the concept of nonmaleficence is contained within the concept of justice (it is incoherent to suppose that one could be both just and cruel),³ the theory would be awkward at best if it presumed to explain the latter and not the former.⁴ Nonetheless, Rawls clearly sees tension between contractarianism and one's presumed duty of nonmaleficence to non-rational creatures.

In a similar vein, David Gauthier writes:

Only beings whose physical and mental capacities are either roughly equal or mutually complementary can expect to find cooperation beneficial to all. Humans benefit from their interaction with horses, but they do not cooperate with horses and may not benefit them.... We may condemn all coercive relationships, but only within the context of mutual benefit can our condemnation appeal to a rationally grounded morality.⁵

Aside from the odd comment that we "may not"—cannot, might not, ordinarily do not?—benefit horses (by giving a horse open pasture, sugar cubes, and protection, do we not benefit it?), Gauthier makes it clear that a "rationally grounded morality" cannot

explain why it is immoral to exploit animals strictly for our own purposes and possibly to their detriment.

2. Three Possible Solutions

The most obvious solution to this dilemma is simply to grasp one of its horns. I do not intend to offer any sustained defense or refutation of contractarian theories of moral obligation, nor would the scope of this paper allow it. Let us assume, therefore, *for the sake of the argument*, that some version of contractarianism is true, thereby eliminating one solution to the problem.

This leaves, of course, the other option, where we deny that we have any direct moral obligations to animals and severely defective humans. Doing this, however, runs counter, at least in the latter instance, to some of our most deeply held moral convictions. If we accept that we have no direct duties to the severely retarded, then, as Peter Singer points out, we are committed to "holding that mental defectives do not have a right to life and therefore might be killed for food...or for the purpose of scientific experimentation."⁶ This is indeed an unsavory position and as such ought to be avoided as long as rational alternatives exist.

Intuitive appeals of this sort have lost favor among philosophers in recent years, but fortunately we need not rely solely on intuition to address the point. There are sound theoretical reasons for claiming that severely defective humans, at least, are appropriate objects of direct moral concern. The simple fact that they have awareness of their surroundings, some sense of continuity of self through time, and an ability to experience pleasure and pain are all reasons sufficient to explain their membership in the moral community. They are subjects of biographical lives, and because of this we have moral duties toward them.

Another possible solution lies in rescuing our intuitions by separating animals from severely defective humans. If there is a significant moral difference between them, then we could accept contractarianism and avoid the counter-intuitive results imagined above. Of course, this option requires at least two things. First, it must be explained how human yet nonrational creatures are included within the social contract. Second, it must be explained how the first explanation applies to severely defective humans but not to at least some animals. In other words, what possible *moral* difference exists between them?

Neither of these tasks is easy. I shall postpone a discussion of the first approach, but it is fairly clear that even if we can account for direct moral obligations to defective humans under contractarianism, we cannot do so in such a way that excludes nonhuman animals. What possible morally relevant difference exists between them? In the case of fetuses or newborns, we might appeal to their *potential* rationality, but this response, although excluding animals, does not address the severely retarded whose potentialities are quite limited. The only other candidates are an appeal to religious doctrine or species membership. Although the former allows for a real moral distinction, it is doubtfully true.⁷ The latter is certainly true, but allows for no moral distinction. Several authors have attempted to explain why species membership *simpliciter* is morally significant, but none of the attempts succeed.⁸ They all employ one of two tactics. Either one sees species membership as morally significant because it represents a family relationship that persons share with the less fortunate members of the species, or some species are deemed morally significant because their members *typically* share certain morally important properties. On either understanding, we owe other humans more moral consideration than we do members of other species *regardless of any other properties each may possess*.

Understanding moral obligation in terms of family relationships poses a difficulty for any attempt to specify just which family groupings are significant. The defenders of this approach have species in mind, but one could repeat the same reasoning and insist that we ought to treat members of our own race better than members of other races due to the family resemblance of skin color. This type of argument could be repeated for a potentially infinite number of classifications. Note that a defender may not argue that some classifications have moral relevance and others do not because, according to this approach, we assign moral worth only *after* we have decided upon a classification.

The second approach recognizes that the possession of various morally relevant properties (sentience, rationality, etc.) is what explains moral worth, but in an oblique way. This perspective recognizes the moral value of beings who belong to *any* classification whose members typically possess the necessary properties. This theory of moral worth avoids the problem encountered by the first attempt, but it runs afoul of an entirely different problem. Michael Tooley illustrates the difficulty:

Let John and Mary be two individuals that are not rational beings, and that belong to different species, but which are indistinguishable with respect to their psychological capacities and their mental lives. It is possible that John belongs to a species 99% of which are rational beings, while Mary belongs to a species of which only 1% are rational beings. If species membership is morally relevant..., it will be wrong to kill John, but may very well not be wrong to kill Mary....⁹

Tooley's far-fetched thought experiment reveals a purely theoretical objection to the theory of moral obligation in question. Any attempt to define the moral community according, not to the capacities that individual prospective members possess, but, rather, according to the groups to which they belong, is unsatisfactory. If we start down the road of including beings within our scope of moral concern simply because they are members of some group or another, we may just as easily *exclude* others because of their group membership. This strategy is, of course, typical of racism, sexism and all of the other 'isms' which seek to assign individuals less moral value than others based on their membership in some group.

One might object, noting that species membership is merely a sufficient, rather than a necessary, condition for inclusion within the moral community; hence, we do not risk this type of exclusion. However, this response misses the point. The issue concerns how mere species membership serves *any* normative role in our moral reasoning, and adopting one explanation over another fails to account for how species membership is relevant in the first place.

There is, of course, at least one sense in which an appeal to group membership is appropriate in this context. There is nothing at all objectionable in recognizing that we have moral duties to all and only those beings who are members of the group defined as "those possessing morally relevant properties." But this understanding, although unobjectionable, completely begs the question. Group membership may always be appealed to in the trivial sense that those possessing morally relevant properties can be recognized as constituting a group. The difficulty lies in determining which properties are morally relevant, and I know of no remotely plausible explanation for the moral relevance of species membership.¹⁰

The upshot is that if we wish to maintain that defective humans have moral value and deserve our protection, then we are rationally bound to take the same position relative to those nonhuman animals whose cognitive capacities are at a similar or greater level of development. All that remains is a strictly empirical matter of determining which animals possess the necessary capacities, a task better suited to cognitive psychology than moral philosophy.

3. A Promising Approach

One model suggests a possible solution. It is common for one party to enter into a contract with another for the purpose of serving, not one's own welfare, but the welfare of a third party, such as when parents establish a contract with a day care center for the purpose of serving a child's interest. This type of situation is quite common and suggests a promising approach to the current problem. Perhaps we ought to think of animals and severely defective humans as appropriate objects of direct moral duty much as we think of children in this way. Young children lack the capacity to make agreements, and yet rational beings regularly assume contractual obligations toward them at the request of other rational beings with an understandable interest in their own children's welfare.

In the example above, the day care operator enters into the contract solely for the purpose of serving her own interests. The parents' reasons are perhaps, although easy to understand, somewhat difficult to explain.

Two explanations seem plausible. Either the parents identify their own interests with that of the child's so intimately that their motive is, in an indirect fashion, also based in self-interest, or natural parental affection motivates a spirit of altruism. According to the former explanation, parents are self-interested. The latter explanation suggests that they are self-sacrificing. Whichever explanation we accept, the parental model fails to explain how, according to contractarianism, we have direct moral obligations to nonrational beings.

Let's consider each account in turn. According to the self-interest model, we must imagine that rational beings would identify with mental defectives and animals. Why would this be true? Of course, in those cases in which an ordinary adult human has a *special* relationship with either—as in the case of a pet or a severely retarded child—the self-interest model might

apply. However, contractarian morality tries to provide a *general* explanation of moral obligation that does not depend on the presence of any specific prior affections. Contractarian morality asks us to assume that, while signing the hypothetical contract, one has no particular friends or family to consider. We have only rational self-interest, and from this perspective it is radically unclear why one would closely identify one's own interest with the interests of nonrational creatures and thus have any moral obligations toward them.

On the other hand, the altruistic account offers a ready explanation for including the nonrational within the scope of moral concern. We care for animals and defective humans simply out of a spirit of compassion. This account provides the needed explanation to be sure, but it is one clearly inconsistent with contractarianism. Contractarian morality is intentionally heartless. It attempts to explain moral obligation within a purely self-interested, nonaltruistic framework.

The parental model thus fails. Either it fails to provide the required moral explanation, or it does so at the expense of the very theory it purports to exemplify. It seems, therefore, that if we wish to account for moral obligations owed to the nonrational by the rational within a contractarian framework, we must find an explanation elsewhere.

4. A Solution

Having ruled out a rejection of contractarianism, an attempt to deny all direct moral duties to animals and defective humans, an attempt to affirm moral duties to defective humans while denying the same to nonhuman animals, and a parental model of contractarian morality, the only remaining alternative lies in a compatibilist position explaining how contractarianism might account for an inclusion of both within the moral community.

As stated at the outset, most commentators on the subject have rejected this approach, preferring instead to emphasize incompatibility. To appreciate the strength of this position, we should review the reasons that motivate it. Contractarianism seems to exclude non-rational creatures from the moral community because, according to the theory, the moral community is *defined* in terms of those who participate in the contract. This is understandable. According to the political model, individuals come together out of self-interest and live under the terms of a contract. Society includes only those who explicitly or implicitly agree to the terms of

the contract. Similarly, on the moral model, individuals agree to live under a set of moral rules comprising the entirety of the rights and obligations to which persons are subject. The moral community is comprised of only those who, in a hypothetical circumstance to be specified, agree to its terms. Since only rational beings can make agreements, only rational beings are subject to and protected by the rules that result.

If a contractarian account of moral obligation includes animals and defective humans, then it must not follow too closely in the above respect its political analogue. There are, in any case, good reasons not to understand a moral theory strictly in terms of a political model. A political, social contract necessarily includes rather specific facts concerning human nature and the natural surroundings. The simple fact that the globe includes scores of independent political systems testifies to the importance of ethnic, geographical, and historical facts in binding us together in a bewildering number of ways. According to the moral model, though, such contingencies are presumably less important. Indeed, Rawls' account in particular asks us to imagine a hypothetical original position in which such facts are not known. The moral rules that result exist on a much higher level of abstraction than corresponding political rules of the traditional social contract.

Appreciating this relatively high level of abstraction helps explain how the interests of nonrational beings can be included within a Rawlsian-type contract. Those in the original position are rational and self-interested, but there is no guarantee that their status remains so after emergence from behind the veil of ignorance. The capacities necessary for rational choice possessed by those behind the veil of ignorance need not endure for all after incarnation. We require only that the beings to whom we owe justice are such that they may *benefit* from the rational, self-interested choices made in the original position. If the veil of ignorance is employed to ensure that choices concerning a future account of justice do not discriminate against particular social, racial, or economic groups who stand to benefit from just treatment, then it may also obscure the possibility of incarnation as a mental defective or animal. The one restriction on the scope of this incarnation would be that only those who may benefit from just treatment are considered (it would be senseless to speak of treating a rock or a bug unjustly).¹¹

Quite apart from Rawls' specific proposal, we can approach the issue from the general perspective of the

nature of contracts and their results. The social contract theory argues against including non-rational creatures within the scope of the moral community since the benefits of the contract are designed to accrue to only those who agree to it. This is because the contract is motivated out of self-interest, and all of the rules that result prescribe behavior which serves the collective long-term interest of the contractors.

The moral rules that result from a hypothetical contract, while motivated out of self-interest, must include elements such as justice and fairness whose application does not in *all* cases maximize self-interest for all the contractors. In other words, applying such principles may require that persons occasionally forgo their own interests in recognition of the interests of others. Indeed, Gauthier sees that the great challenge of formulating a rationally grounded morality lies in defending the seemingly paradoxical claim that pursuit of a long-term strategy of self-interest often requires that one strikes a bargain that is less advantageous than some alternative. Contractarian morality has excluded nonrational creatures from the moral community because it limits the benefits of the contract to only those who are at least capable of participating in its formulation. This assumption is captured by the passage from Gauthier cited earlier in which he notes the impossibility of rational cooperation between horses and humans.

So, if contractarian morality allows, even demands each contractor to forgo maximum benefit to one's self, all that remains is the question of whose interests the theory allows us to consider. As a question about who may *benefit* from a contract, rather than who can *participate* in one, the traditional argument that non-rational creatures are not, indeed, cannot, be covered by the contract seems odd.

For instance, spectator sports often seem at least partially designed to benefit non-players—those not even capable of playing!—as well as players. So, too, it is not unreasonable that the moral rules resulting from the social contract may be designed to benefit those who lack moral agency and who cannot, following the analogy, play the game.

The goal of any theoretical account of moral obligation is to determine not merely the moral status of this or that action but of whole *classes* of actions; for this reason, a satisfactory moral theory must be stated at a relatively high level of abstraction. For contractarianism this has meant a focus not on any existing contract but, rather, on a hypothetical contract that

rational persons would agree to under an imaginary set of circumstances. After all, human beings almost certainly never explicitly agreed to any contract outlining the set of all of our moral rights and duties.

This requirement, demanding that we look to hypothetical, not actual, contracts, provides a theoretical explanation for how the welfare of the nonrational can be provided for under a contractarian framework. All we need ask is what *would* a nonrational being bargain for if it were temporarily able? The fact that the ability to bargain rules out any actual nonrational creature is beside the point, since the nonrational creature imagined is rational only while making the agreement. This is not as farfetched as it may seem, since the temporarily rational regularly make agreements similar to this—as, for example, when one leaves instructions in a living will directing how one wants to be treated if through disease or accident one becomes nonrational.¹²

Similar to this, we can imagine a (obviously hypothetical) temporarily rational horse bargaining for the sorts of advantages that would benefit ordinary horses. Since ordinary horses have relatively limited capacities to appreciate any supposed advantage, the moral rules that result would be necessarily one-sided. Humans would have real duties to horses, but their duties to other humans would be considerably more stringent and comprehensive.

There is yet another explanation for how the scope of a contract can be supposed to include those other than the initial contractors. It is not surprising that rational, self-interested beings often enter into contracts intending to benefit others as well as themselves. The reasons for this are psychological and varied, but perhaps the most obvious is that persons typically prefer engaging in activities that allow for a sense of a shared experience. We do not need to assume parental altruism to recognize that we often engage in social behavior precisely because, all else being equal (like the man who refuses to drink alone or the artists and athletes who prefer to exercise their gifts for the benefit of an appreciative audience), we prefer sharing our well-being with others. If we make this psychological assumption, we see that rational agents may opt for the sort of contract which provides the greatest benefit to all those *capable* of benefiting from it. Extending the benefits of one's good fortune to others, so long as one does not as a consequence seriously undercut one's own reward, is, at a minimum, quite consistent with rational decision-making.

We require further study to determine with any precision how to balance, according to contractarian morality, our own interests against the welfare of the nonrational. In the case of animals, it may turn out that depriving ourselves of the taste of their flesh is a relatively minor sacrifice in contrast to the benefits ceded to them. On the other hand, and more likely, even on the contractarian account, we are obliged to treat animals much better than we already do. With respect to seriously defective humans, my sense is that our treatment of them is already quite decent. The above analysis does not recommend a difference in treatment of them; instead, it offers an alternative theoretical explanation that replaces mere species-preference.

So long as we recognize that animals and seriously defective humans are similar to us in the morally relevant sense that they, like us, have an interest in avoiding pain and engaging in pleasurable activities in a stimulating environment, we can, according to the analysis above, explain moral duties toward them.

By way of a conclusion, in the first instance contracts may issue from Rawlsian hypothetical circumstances which include the possibility that the contractors will be incarnated as nonrational beings. If so, then the rules that are agreed on ought to speak to the interests of those who are capable of *benefiting* from decent treatment as well as those who are capable of initiating it. Second, *any* plausible contractarian account of morality conceives of the contract as hypothetical, not actual, and there is no obvious reason why the hypothetical circumstance conceived may not be framed in terms that a creature would demand for itself if it were temporarily capable of rational activity. Third, human beings typically prefer agreements that allow for a wider, rather than a narrower, range of benefit out of a natural desire to share the benefits of one's own good fortune so long as this sharing does not entail a serious diminution of one's own welfare.

If this analysis is correct, then we ought no longer test the plausibility of a contractarian theory of morality against implications it has for seriously defective humans. Also, the price we pay for aligning contractarian theory with intuitions concerning our duties to defective humans forces us to accept similar duties toward some animals. This latter conclusion, although not strongly supported by intuition, is an inevitable consequence of other intuitions that most of us are reluctant to abandon.

Notes

¹ John Rawls, *A Theory of Justice* (Harvard University Press, 1971); and, David Gauthier, *Morals By Agreement* (Oxford University Press, 1986).

² Rawls, p. 512.

³ The U. S. Constitution recognizes this. Allowing "cruel and unusual punishment" is simply inconsistent with a civilized rendering of criminal justice.

⁴ Rawls adds to the confusion by identifying both the "natural duty not to be cruel" and the "duty of justice" as natural duties, thereby reinforcing the notion that both are basic moral notions treated by his theory.

For an extended criticism of Rawls' theory as it applies to animals see, Tom Regan, *The Case for Animal Rights* (University of California Press, 1983), pp. 163-74.

⁵ Gauthier, p. 17.

⁶ Peter Singer, "Animals and the Value of Life," in *Matters of Life and Death*, edited by Tom Regan (Random House, 1986), p. 366.

⁷ For instance, the doctrine of dominion roughly states that God gave man dominion over all the creatures of the earth. Humanity's superior status is justified by the possession of an eternal soul (in the image of God), which animals lack. As Singer points out (*op. cit.*, pp. 346-48), even if we accept that humans have an immortal soul that animals lack, it is unclear why this argues for rather than against the permissibility of taking animal life.

⁸ See, Jean Beer Blumenfeld, "Abortion and the Human Brain," *Philosophical Studies*, Vol. 32 (1977): pp. 251-68; Carl Cohen, "The Case for the Use of Animals in Biomedical Research," *The New England Journal of Medicine*, Vol. 315 (1986): pp. 865-70; and Robert Nozick, "About Mammals and People," *New York Times Book Review*, (November 27, 1983).

⁹ Michael Tooley, *Abortion and Infanticide* (Oxford University Press, 1983), p. 71.

¹⁰ It is necessary to explain this in light of the fact that undoubtedly a vast majority of people would deny it. We tend to think that species membership is morally relevant simply because human beings are the only moral agents of which we are aware. This is, however, merely a contingent fact about our experience.

¹¹ Donald VanDeVeer ["Of Beasts, Persons, and the Original Position," *Monist*, Vol. 63, no. 3 (1979): pp. 372-73] makes a similar point. He writes:

If, then, the original position were fully neutral, its participants would not only have to be ignorant of the race, sex, or social position of its participants

in the just society, it would seem that they would have to be ignorant of their species membership as well—subject only to the qualification that they shall have interests as participants in the just society.

¹² An example from fiction may be helpful here. In "Flowers for Algernon," a severely retarded man is transformed into a genius as a result of taking a miracle drug, only to gradually descend back into imbecility as the effects of the drug wear off. Imagine the sort of demands such a person might make while he was temporarily rational regarding how he wanted to be treated in the future.

